Neurocognition after pediatric heart surgery: a systematic review & meta-analysis

Caroline Sterken[1], MPsych, Jurgen Lemiere[2,3], MPsych, PhD, Ilse Vanhorebeek[1], MEng, PhD, Greet Van den Berghe[1], MD, PhD, Dieter Mesotten[1], MD, PhD

Affiliations: [1]Clinical Department and Laboratory of Intensive Care Medicine, Division of Cellular and Molecular Medicine, KU Leuven, Belgium; [2]Child and Adolescent Psychiatry, University Hospitals Leuven, Belgium, [3]Paediatric Haemato-Oncology, University Hospitals Leuven, Belgium

Address correspondence to: Dieter Mesotten, Dept Intensive Care Medicine, KU Leuven, Herestraat 49, B-3000 Leuven, Belgium, [dieter.mesotten@med.kuleuven.be], telephone number +32 16 344021, fax number +32 16 344015

Short title: Neurocognition in pediatric heart patients

Abbreviations: ASD = atrial septal defect; Att = Attention; CHD = congenital heart disease; CI = confidence interval; CPB = cardiopulmonary bypass; ECMO = extracorporeal membrane oxygenation; EF = executive function; IQ = intelligence quotient; IQR = interquartile range; Mem = memory; NEPSY = NEuroPSYchological Assessment; SD = standard deviation; SMD = standardized mean difference; LGC = Leuven glucose control; TOF = tetralogy of Fallot; VSD = ventricular septal defect

Key Words: congenital heart disease, cognition, attention, executive function, intelligence, children

Clinical Trial Registration: clinicaltrials.gov Identifier NCT00214916

Key messages

What is already known about this subject?

Children with congenital heart disease risk difficulties in academic and daily life functioning, which are linked to intelligence and specific neurocognitive skills such as executive functions, attention and memory.

What does this study add?

This meta-analysis demonstrates that children with congenital heart disease who underwent heart surgery, show impairment in basic alertness, which is essential for intellectual, executive and academic functioning.

How might this impact on clinical practice?

Physicians need to be aware that additional testing for alertness may be helpful for children with CHD who experience subtle, but pervasive difficulties in daily and academic functioning,

Abstract

Background and objective: Children with congenital heart disease (CHD) often experience difficulties in academic and daily functioning, which have been associated with intelligence and neurocognitive skills, like executive functions (EFs), attention and memory. We report the neurocognitive data of children with CHD who were included in the Leuven glucose control trial (LGC-trial). Through a systematic review and meta-analysis, we aimed to find which neurocognitive functions are most consistently and prominently affected.

Methods: 365 children with CHD and 216 healthy control children underwent extensive neurocognitive testing in the LGC-trial. A comprehensive search of electronic databases PubMed, Embase and Cochrane was conducted for studies measuring intelligence, EFs, attention and memory in children who underwent heart surgery for CHD. Standardized mean differences (SMDs) between the CHD-group and a healthy control group were calculated for these neurocognitive functions. LGC-trial data were included in the meta-analysis.

Results: 12 studies with a healthy control group were included in the meta-analysis, involving 647 CHD-patients and 633 controls. The CHD-group (median age 7.35 years at testing) scored worse than healthy control children for all investigated neurocognitive functions. A medium SMD was found for intelligence (SMD=-0.53 [95% CI: -0.68 to -0.38] $p<0.00001$). Alertness, an attentional function, was also consistently poorer in the CHD-group. Memory was less affected, while EF had a medium SMD with large heterogeneity.

Conclusions: Children with CHD risk lower performance on intelligence and alertness, which may contribute to daily life and school difficulties. Heterogeneity in neurocognitive assessment and small sizes in most studies limit the interpretation.

Introduction

As the survival rate of children with congenital heart disease (CHD) continues to improve thanks to medical advances,[1] public interest and research have been focusing more on how children with CHD survive.[2] Children with CHD may experience difficulties in daily[3] and academic functioning[4], which may persist into adulthood.[5] In this systematic review and meta-analysis we aim to examine intelligence and specific neurocognitive skills, especially executive functions, as important predictors for academic and daily functioning in children with CHD after heart surgery.

Intelligence scores (IQ) are a good predictor of academic performance.[6] However, children with CHD often have intelligence scores within average range compared to population norms.[7] Research in other pediatric conditions, such as traumatic brain injury, has demonstrated that intelligence scores are rough measures, which fall short to detect more specific neurocognitive skills like executive functions (EFs).[8] EFs cover a variety of cognitive functions, such as planning, organization, flexibility, cognitive control and working memory. These are essential in many domains of daily life[9] and contribute to academic performance.[10,11] Other neurocognitive skills like memory and attention, which are interrelated with EFs, also contribute to academic performance. [9]

There is growing evidence that brain development of children with CHD can differ from normal brain development. This misdevelopment may even start prenatally due to impaired cerebral blood flow.[12] Postnatally, infants and older children with CHD may have preoperative[13] and postoperative white matter abnormalities,[14] which may relate to worse neurocognitive outcome in children with CHD.[15] Surgery seems to impact neurocognitive outcome in CHD as well.[15] Considering the vulnerability of their brain, neurocognitive functions may be worse in children with CHD than in healthy control children without CHD.

Some relevant reviews have discussed the importance of these neurocognitive functions.[2,16,17] However, to our knowledge, the outcome data of specific neurocognitive functions have not been analyzed systematically in children with CHD. One systematic review[18] and one meta-analysis[19] examined intellectual outcome but not more specific neurocognitive skills in children with CHD after heart surgery.

A large randomized controlled trial (n=700) in which neurocognitive development of children was assessed 4 years after critical illness and treatment with tight glucose control has recently been completed.[20] The results demonstrated that tight glucose control in critically ill children improved motor coordination and cognitive flexibility in comparison with children in whom blood glucose levels up to 215 mg/dL were tolerated.[20] Seventy-five % of the study population in this Leuven glucose trial (LGC-trial) underwent heart surgery for congenital heart defects. Neurocognitive data of the heart surgery subgroup have not been analyzed yet.

Thus, the first aim of this paper was to report the neurocognitive data of this large cohort of children with heart surgery for CHD, included in the LGC-trial, and healthy controls. The second aim was to do a systematic review and meta-analysis for intelligence, EFs, attention and memory in children with CHD after heart surgery. We hypothesized that specific neurocognitive skills like EFs are more impaired than intelligence in children with CHD.

Methods

1) Analysis of the LGC-trial data

Data of all children who underwent neurocognitive testing and for whom a full-scale IQ was available, were analyzed (children with CHD n=361, healthy controls n=215). For four children of the CHD-group and one child of the control group, a full-scale IQ could not be calculated. Baseline neurocognitive data, when the patients were included into the LGC-trial, were not available. Demographic, clinical and neurocognitive data are reported as numbers and percentages, or, median and interquartile range (IQR). Because of imbalance between the CHD- and control group for gender, the presence of a syndrome, socio-economic status and age at follow-up, propensity score matching was performed using IBM SPSS Statistics 22.0.0.1 and R statistical software version 2.15.3. For more details on propensity score matching, we refer to **eMethods1**. Demographic, clinical and neurocognitive data of the tested and matched population were further analyzed using chi-square test for dichotomous variables and unpaired non-parametric Wilcoxon rank-sum tests for continuous variables (JMP version 11.2.0 [SAS Institute]). The details of the study protocol and neurocognitive test battery have previously been reported.[20]

2) Systematic review & meta-analysis

Data sources and searches

A comprehensive search of electronic databases PubMed, Embase and Cochrane was conducted for studies published between the beginning of each database and December 2014. Each search strategy consisted of four major parts: cognition, heart, child and tests. Both index language terms (MeSH, Emtree) and keywords were used in every part of the search

strategy (See **eMethods 2**). We also manually screened reference lists of studies identified through database search.

Study selection

Two selection criteria were premised for title and abstract screening. First, the study population consisted of infants, children, adolescents and/or young adults (<24 years old) with CHD. Second, they needed to have at least an intelligence testing with an indication of a specific measure of EF, memory or attention, or broader neurocognitive assessment. For the full-text screening we introduced two more selection criteria. First, the subjects had neurocognitive testing after heart surgery or an interventional cardiac procedure. Studies that tested subjects in the first six postoperative months were excluded to avoid interference with the acute medical phase. Second, at least one test that measured EF, memory and/or attention was needed.

Data extraction

First, we gathered data about the sample size, the age at surgery and the age at testing of the tested groups. We also investigated whether the same study population was tested in already included studies of the same research group. Furthermore, for the comparison of the CHD-group versus a healthy control group, we collected data necessary to quantify differences between the CHD- and healthy control group. The collected data are summarized in **Table 1**. In case of missing data for the quantitative analysis, authors were contacted.

Data analysis

As appropriate, mean (standard deviation) or median (interquartile range) sample size, age at surgery and age at testing were calculated for all studies. Risk of bias of all included studies [20-31] was assessed by two raters independently (CS and JL) by means of a modified version of

the Downs and Black checklist.[32] Six items that focus on RCT's (items 4, 8, 14, 19, 23, 24), were omitted, leading to a maximum total score of 22. In case of different scores by the two raters a consensus was reached through discussion. Inter-rater reliability was good (Spearman's Rho 0.722; p=0.008). Depending on the number of included studies, mean (standard deviation) or median (interquartile range) total risk of bias and sub-scores were calculated for all studies.

Analyses were performed on neurocognitive data of studies that used a healthy control group. To combine these continuous data measured by different instruments, effect sizes, i.e. standardized mean differences (SMDs), were calculated in Review Manager 5.2 by means of Inverse Variance, random effects analysis (http://ims.cochrane.org/revman). SMDs were calculated for the following neurocognitive functions: intelligence, alertness (attentional function), memory, and inhibition (executive function). Neurocognitive data of the LGC-trial were included in the SMD-calculation. A two-sided p-value <0.05 was considered statistically significant. SMDs were classified according to Cohen's guidelines: d = 0.2 defined as small, d = 0.5 as medium and d = 0.8 as large.[33] For clinical interpretation, the overall SMD for intelligence scores was multiplied by the typical standard deviation (SD) of the normal IQ distribution (mean 100 ± SD 15). The I^2-statistic was used to evaluate heterogeneity. $I^2 \geq 50\%$ was defined as substantial heterogeneity. Funnel plots were also made in Review Manager, when at least 10 studies were available, to explore small-study effects.

For more details on study selection and data analysis we refer to **eMethods 1**.

Results

1) Analysis of the LGC-trial data

Demographic and clinical data of tested and matched post-heart surgery population and healthy controls are presented in **Table 2**. Neurocognitive data are presented in **Table 3**. Children of the tested post-heart surgery population score worse for intelligence, visual-motor integration and all measures of alertness, motor coordination, inhibition, flexibility (except for ΔNo. of errors), memory and behavior, compared to healthy controls. After propensity score matching, the CHD-group scores worse for intelligence, visual-motor integration, motor coordination (alternating taps), inhibition (Δ No. of errors), memory (verbal working memory and immediate memory) and behavior (internalizing and total problems).

2) Systematic review & meta-analysis

Flow diagram

Figure 1 shows the article screening phases and the reasons for exclusion, according to the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-analysis).[34] An overview of the 12 included studies is presented in **eTable 1.**

Sample size, age at surgery and at testing

The median sample size of the CHD-groups in all 12 studies that were included in the meta-analysis was 31 (IQR 18-43) The median sample size of the healthy control groups was 33 (IQR 20-42). The median age at surgery was 1.25 years (y) (IQR 0.0.7-3.85y) and the median age at testing was 7.35 y (IQR 5.70-8.37).

Quality assessment

The median (interquartile range, IQR) risk of bias of all 12 included studies was 16 (IQR 15-17). Median score for reporting was 8 (IQR 7-8) with a maximum score of 9. Median score for external validity was 2 (IQR 1-2) with a maximum score of 3 and median score for internal validity was 7 (IQR 5-7) with a maximum score of 9. Only 1 out of 12 studies mentioned that the study had sufficient power to detect a clinically relevant difference.

Meta-analyses: CHD-group versus healthy controls

Total SMDs for all neurocognitive domains indicate lower scores for the CHD-group. A medium SMD with low heterogeneity was found for intelligence (SMD=-0.53 [95% CI: -0.68 to -0.38] p<0.00001, I^2=32%) (**Figure 2**), indicating a drop of 0.53 times the standard deviation of the normal IQ distribution: -8 IQ points. A medium SMD was also found for alertness non-reaction time (SMD=-0.47 [95% CI: -0.67 to -0.27] p<0.00001, I^2=19%) (**Figure 3**) and a smaller SMD for alertness reaction time (SMD=0.25 [95% CI: 0.08 to 0.42] p=0.004, I^2=0%) (**eFigure 1)**. Verbal memory showed a smaller SMD (SMD=-0.35 [95% CI: -0.54 to -0.15] p=0.0004, I^2=0%) (**Figure 4**), while non-verbal memory did not differ between children with CHD and healthy controls **(eFigure 2)**. EF reaction time, examining the inhibition function, had the largest SMD, but a high level of heterogeneity (SMD=0.57 [95% CI: 0.10 to 1.04] p=0.02, I^2=80%) (**Figure 5**). EF non-reaction time for inhibition also showed a medium SMD but with low heterogeneity (SMD=-0.51 [95% CI: -0.74 to -0.29] p<0.00001, I^2=0%) (**eFigure 3**). The funnel plot for intelligence was slightly asymmetrical (**eFigure 4**). Due to the limited availability of studies on the other neurocognitive functions, funnel plots could not be built for alertness, memory and inhibition.

Discussion

This meta-analysis and the results of the LGC-trial provide sound evidence that not only intelligence but also more specific neurocognitive functions are impaired in children with CHD who underwent heart surgery, compared with healthy controls. Contrary to our hypothesis, intelligence, EFs and alertness seem to be equally affected.

Intelligence

The SMD of -0.53 (95% CI -0.68, -0.38) in the meta-analysis and the effect size of 8 IQ-points in both the meta-analysis and LGC-trial are of similar magnitude to these in prematurely born children[35] and are in line with the meta-analysis of Karsdorp et al.[19] Despite the worse intelligence scores for the CHD-group compared to healthy controls, the mean IQ scores of the children with CHD were still within the "normal" average range, when comparing with population norms. Nevertheless, these IQ differences can have consequences in a learning environment.[36] Population norms might insufficiently capture the dynamic nature of development. A well-matched healthy control group of typically developing children without CHD may offer a more representative reflection of normal variation between and within typically developing children[37,38] who grow up in the same time period as children with CHD.

Alertness, inhibition and memory

Similarly to intelligence, alertness was also consistently impaired across the available studies in the meta-analysis. The alerting system is one of the attentional networks in addition to the orienting and executive attention network.[39] It is responsible for achieving and maintaining a state of high sensititivity to incoming information. An efficient alerting system is pivotal for

other more complex cognitive functions, such as EF. Additionally, detecting problems of alertness has an important clinical consequence because improving alertness is considered essential for cognitive rehabilitation.

Recently, more consideration is given to impairments in EF in children with CHD.[2,16,17] The LGC-trial showed that in critically ill children exactly the more complex EF, flexibility, was improved by tight glucose control, almost to the levels of healthy controls[20] However, from this meta-analysis we have to conclude that studies examining these specific neurocognitive skills are scarce. The few studies included in this meta-analysis involved several specific subgroups of CHD with only a limited number of patients in each subgroup. This resulted in low statistical power and precision. Also the use of different tests for EF, adding to the ascertained heterogeneity, may have had an impact on low precision.

Nevertheless, the SMD (0.57) for EF reaction time was in line with the deficit in intelligence (0.53). It indicates that children with CHD react slower when performing specific EF-tasks measuring inhibition. Children with CHD from the LGC-trial showed the smallest SMD in EF reaction time.[20] This may be attributed to the larger size of the study and to the fact that the EF reaction time was corrected for baseline speed and the propensity score matching. The CHD-group of the LGC-trial also made more errors during an inhibition test. This meta-analysis confirmed the findings from previous reviews[2,16,17] that the risk of memory deficits in children with CHD may be lower. However, results from the LGC-trial showed worse performance for immediate memory and working memory of verbal information in the CHD-group. The use of pooled memory scores in the meta-analysis may have hidden specific memory deficits. Children may also have been too young to detect differences in tests, which are examining functions that are continuously developing through childhood.

The aforementioned indicates that the difficulties, which children with CHD experience when performing IQ-tests and complex tasks requiring EFs, may be explained by a basic alertness

deficit. This has also been found in other pediatric populations and may be linked with white matter changes.[40] Alternatively, other EFs like flexibility[20] or working memory may preferentially be affected.

Strenghts

This systematic review and meta-analysis has several strengths. Exploring the effects of CHD on not only intelligence but also EFs, attention and memory in the same patients has been a new avenue in the assessment of neurocognitive function in children with CHD. This allowed us to compare the impact of CHD on intelligence with the impact of CHD on specific neurocognitive skills. The studies which included a healthy control group could be combined for the analysis of a SMD, despite the use of different tests. Therefore, the findings are fairly robust. The analysis of the LGC-trial data offers a valuable contribution of neurocognitive data of a large CHD- and healthy control group, thereby increasing power and precision in the meta-analysis. Certainly, propensity score matching for relevant factors in neurocognitive development improved the stringency and reliability of the analyses.

Limitations

The meta-analysis has a few inherent limitations though. First, attention and EFs involve several functions, which may interact.[9] Although the lower aggregated score on intelligence is clinically relevant, it is not clear whether the poorer results for attention and EF have any clinical impact, because universal definitions and test protocols are lacking. Unfortunately, the number of studies assessing different aspects of attention and EF was scarce. As a result, only one attentional function and one EF could be examined in this meta-analysis. Therefore, no conclusions can be drawn for other attention components and EFs. Future studies thus ought to use comparable test batteries for neurocognitive function assessments. Second, the

very small sample sizes of the CHD- and control groups, partially due to the separate reporting of a high variation in CHD diagnoses, reduced statistical power and precision both at the individual study level and at meta-analysis level. Because of the separate reporting of CHD-subgroups, neurocognitive data of healthy controls were sometimes included twice in forest plots. The slightly asymmetrical funnel plot of intelligence reflects a possible publication bias and may have overestimated the effect size. Future research should pay more attention to the statistical powering of effects on neurocognitive outcome. Third, the exclusion of six non-English studies (one Italian, one Chinese, one French, one Spanish and two German) might have raised a language bias. However, research has shown no evidence of systematic bias from the use of language restrictions in systematic review-based meta-analyses.[41] Third, this meta-analysis cannot draw any conclusion on the interaction between effect of IQ and alertness, EF, and memory. At least to exclude that effects on EFs and memory are not entirely explained by a generalized or more basic impairment, one ought to examine IQ and alertness in addition to the other functions. Fourth, studies on EFs, attention and memory outcome in adolescence are lacking, indicating the need for long-term longitudinal follow-up studies up to secondary education,[18] when academic difficulties can appear. Adolescence is particularly essential for the late maturation of the prefrontal cortex and EFs.[37] Understanding the impact of these prefrontal cortex functions on daily and school functioning is clinically relevant, because a better understanding and early detection of deficits may improve daily functioning of children with CHD.[2]

Conclusions

Children with CHD who underwent heart surgery, consistently perform worse for intelligence and alertness. Memory appears to be less affected from this meta-analysis. The effect of CHD on EF in young children cannot be reliably assessed due to poor standardization of the testing

methodology. Larger, more standardized, longitudinal long-term follow-up studies of specific neurocognitive skills in a large group of children with CHD and a matched healthy control group, are necessary for a better understanding of neurocognitive deficits and their impact on daily life and school functioning.

Acknowledgments

We thank Jens De Groot (Biomedical Library, Catholic University Leuven) for his help with the search strategies.

Contributions

Ms. Sterken conceptualized and designed the study, carried out analyses, drafted initial manuscript and approved the final manuscript as submitted.

Dr. Lemiere conceptualized and designed the study, carried out analyses, drafted initial manuscript and approved the final manuscript as submitted.

Dr. Vanhorebeek carried out analyses, reviewed and approved the final manuscript as submitted.

Dr. Van den Berghe conceptualized and designed the study, carried out analyses, reviewed and approved the final manuscript as submitted.

Dr. Mesotten conceptualized and designed the study, reviewed and revised the manuscript and approved the final manuscript as submitted.

Financial support

Funding Source: The study was funded by the Flemish Agency for Innovation by Science and Technology (IWT/070695/TBM). DM holds a FWO-Flanders Senior Clinical Investigator Fellowship. GVdB via the University of Leuven receives structural research financing via the Methusalem program, funded by the Flemish government and holds an ERC advanced grant from the EU-FP7 (2012-321670).

Financial Disclosure: Nothing to disclose

Conflict of Interest: The authors have no conflicts of interest.

References

1. Moons P, Sluysmans T, De Wolf D, et al. Congenital heart disease in 111 225 births in Belgium: birth prevalence, treatment and survival in the 21st century. *Acta paediatrica (Oslo, Norway : 1992).* 2009;98(3):472-477.
2. Marino BS, Lipkin PH, Newburger JW, et al. Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. *Circulation.* 2012;126(9):1143-1172.
3. McCrindle BW, Williams RV, Mitchell PD, et al. Relationship of patient and medical characteristics to health status in children and adolescents after the Fontan procedure. *Circulation.* 2006;113(8):1123-1129.
4. Shillingford AJ, Glanzman MM, Ittenbach RF, Clancy RR, Gaynor JW, Wernovsky G. Inattention, hyperactivity, and school performance in a population of school-age children with complex congenital heart disease. *Pediatrics.* 2008;121(4):e759-767.
5. Kamphuis M, Ottenkamp J, Vliegen HW, et al. Health related quality of life and health status in adult survivors with previously operated complex congenital heart disease. *Heart (British Cardiac Society).* 2002;87(4):356-362.
6. Hansen BM, Dinesen J, Hoff B, Greisen G. Intelligence in preterm children at four years of age as a predictor of school function: a longitudinal controlled study. *Developmental medicine and child neurology.* 2002;44(8):517-521.
7. Forbess JM, Visconti KJ, Bellinger DC, Jonas RA. Neurodevelopmental outcomes in children after the fontan operation. *Circulation.* 2001;104(12 Suppl 1):I127-132.
8. Chapman SB. Neurocognitive stall: a paradox in long term recovery from pediatric brain injury. . *Brain Injury Professional.* 2007;3(4):10-13.
9. Diamond A. Executive functions. *Annual review of psychology.* 2013;64:135-168.
10. Bull R, Espy KA, Wiebe SA. Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Developmental neuropsychology.* 2008;33(3):205-228.
11. Alloway TP, Alloway RG. Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of experimental child psychology.* 2010;106(1):20-29.
12. Limperopoulos C, Tworetzky W, McElhinney DB, et al. Brain volume and metabolism in fetuses with congenital heart disease: evaluation with quantitative magnetic resonance imaging and spectroscopy. *Circulation.* 2010;121(1):26-33.
13. Miller SP, McQuillen PS, Hamrick S, et al. Abnormal brain development in newborns with congenital heart disease. *The New England journal of medicine.* 2007;357(19):1928-1938.
14. Mahle WT, Tavani F, Zimmerman RA, et al. An MRI study of neurological injury before and after congenital heart surgery. *Circulation.* 2002;106(12 Suppl 1):I109-114.
15. Rivkin MJ, Watson CG, Scoppettuolo LA, et al. Adolescents with D-transposition of the great arteries repaired in early infancy demonstrate reduced white matter microstructure associated with clinical risk factors. *The Journal of thoracic and cardiovascular surgery.* 2013;146(3):543-549 e541.
16. Miatton M, De Wolf D, Francois K, Thiery E, Vingerhoets G. Neurocognitive consequences of surgically corrected congenital heart defects: A review. *Neuropsychology review.* 2006;16(2):65-85.

17. Bellinger DC, Newburger JW. Neuropsychological, psychosocial, and quality-of-life outcomes in children and adolescents with congenital heart disease. *Progress in Pediatric Cardiology.* 2010;29(2):87-92.
18. Snookes SH, Gunn JK, Eldridge BJ, et al. A systematic review of motor and cognitive outcomes after early surgery for congenital heart disease. *Pediatrics.* 2010;125(4):e818-827.
19. Karsdorp PA, Everaerd W, Kindt M, Mulder BJ. Psychological and cognitive functioning in children and adolescents with congenital heart disease: a meta-analysis. *Journal of pediatric psychology.* 2007;32(5):527-541.
20. Mesotten D, Gielen M, Sterken C, et al. Neurocognitive development of children 4 years after critical illness and treatment with tight glucose control: a randomized controlled trial. *JAMA : the journal of the American Medical Association.* 2012;308(16):1641-1650.
21. Calderon J, Bonnet D, Courtin C, Concordet S, Plumet MH, Angeard N. Executive function and theory of mind in school-aged children after neonatal corrective cardiac surgery for transposition of the great arteries. *Developmental medicine and child neurology.* 2010;52(12):1139-1144.
22. Calderon J, Angeard N, Moutier S, Plumet MH, Jambaque I, Bonnet D. Impact of prenatal diagnosis on neurocognitive outcomes in children with transposition of the great arteries. *The Journal of pediatrics.* 2012;161(1):94-98 e91.
23. Hovels-Gurich HH, Seghaye MC, Sigler M, et al. Neurodevelopmental outcome related to cerebral risk factors in children after neonatal arterial switch operation. *The Annals of thoracic surgery.* 2001;71(3):881-888.
24. Hovels-Gurich HH, Konrad K, Skorzenski D, Herpertz-Dahlmann B, Messmer BJ, Seghaye MC. Attentional dysfunction in children after corrective cardiac surgery in infancy. *The Annals of thoracic surgery.* 2007;83(4):1425-1430.
25. Miatton M, De Wolf D, Francois K, Thiery E, Vingerhoets G. Neuropsychological performance in school-aged children with surgically corrected congenital heart disease. *The Journal of pediatrics.* 2007;151(1):73-78, 78 e71.
26. Sarrechia I, De Wolf D, Miatton M, et al. Neurodevelopment and Behavior after Transcatheter versus Surgical Closure of Secundum Type Atrial Septal Defect. *The Journal of pediatrics.* 2015;166(1):31-38 e31.
27. Schaefer C, von Rhein M, Knirsch W, et al. Neurodevelopmental outcome, psychological adjustment, and quality of life in adolescents with congenital heart disease. *Developmental medicine and child neurology.* 2013;55(12):1143-1149.
28. Tindall S, Rothermel RR, Delamater A, Pinsky W, Klein MD. Neuropsychological abilities of children with cardiac disease treated with extracorporeal membrane oxygenation. *Developmental neuropsychology.* 1999;16(1):101-115.
29. Van Der Rijken R, Hulstijn-Dirkmaat G, Kraaimaat F, et al. Open-heart surgery at school age does not affect neurocognitive functioning. *European heart journal.* 2008;29(21):2681-2688.
30. Wells FC, Coghill S, Caplan HL, Lincoln C. Duration of circulatory arrest does influence the psychological development of children after cardiac operation in early life. *The Journal of thoracic and cardiovascular surgery.* 1983;86(6):823-831.
31. Wray J, Pot-Mees C, Zeitlin H, Radley-Smith R, Yacoub M. Cognitive function and behavioural status in paediatric heart and heart-lung transplant recipients: the Harefield experience. *BMJ (Clinical research ed.).* 1994;309(6958):837-841.

32. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of epidemiology and community health.* 1998;52(6):377-384.
33. Cohen J. *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ1988.
34. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ (Clinical research ed.).* 2009;339:b2535.
35. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA : the journal of the American Medical Association.* 2002;288(6):728-737.
36. Prentice DA MD. When Small Effects Are Impressive. *Psychological bulletin.* 1992;112(1): 160-164.
37. Casey BJ, Tottenham N, Liston C, Durston S. Imaging the developing brain: what have we learned about cognitive development? *Trends in cognitive sciences.* 2005;9(3):104-110.
38. Waber DP, De Moor C, Forbes PW, et al. The NIH MRI study of normal brain development: performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. *Journal of the International Neuropsychological Society : JINS.* 2007;13(5):729-746.
39. Posner MI. Measuring alertness. *Annals of the New York Academy of Sciences.* 2008;1129:193-199.
40. Potharst ES, van Wassenaer-Leemhuis AG, Houtzager BA, et al. Perinatal risk factors for neurocognitive impairments in preschool children born very preterm. *Developmental medicine and child neurology.* 2013;55(2):178-184.
41. Morrison A, Polisena J, Husereau D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *International journal of technology assessment in health care.* 2012;28(2):138-144.
42. Jenkins KJ, Gauvreau K, Newburger JW, Spray TL, Moller JH, Iezzoni LI. Consensus-based method for risk adjustment for surgery for congenital heart disease. *The Journal of thoracic and cardiovascular surgery.* 2002;123(1):110-118.
43. Leteurtre S, Martinot A, Duhamel A, et al. Validation of the paediatric logistic organ dysfunction (PELOD) score: prospective, observational, multicentre study. *Lancet.* 2003;362(9379):192-197.

Tables

Table 1. Data collected for quantitative analysis

CHD-group: type of CHD, sample size
Control group: type of control group, sample size
Age surgery, age testing
Overlap study population with another included study of the same research group: yes/no and if yes, decision (inclusion/exclusion)
Intelligence: name of test(s), intelligence scores (mean, SD) of CHD- and healthy control group
Memory: name of test(s), memory scores (mean, SD) of CHD- and healthy control group
Attention: name of test(s), attention scores (mean, SD) of CHD- and healthy control group
Executive functions: name of test(s), executive function scores (mean, SD) of CHD- and healthy control group

Abbreviations: SD = standard deviation

Table 2: Demographic and clinical data of tested and propensity score matched post-heart surgery population and healthy controls in the LGC-trial

	Tested population			Propensity score matched population		
	Tested-Post-Heart surgery population (N=361)	Healthy controls (N=215)	p-value	Tested-Post-Heart surgery population (N=167)	Healthy controls (N=167)	p-value
Caucasian race[1]	343 (95.01)	211 (98.14)	0.058	162 (97.01)	164 (98.20)	0.474
Exclusive European[1]	330 (91.41)	201 (93.49)	0.369	160 (95.81)	156 (93.41)	0.332
Exclusive Dutch language[1]	296 (81.99)	186 (86.51)	0.155	144 (86.23)	142 (85.03)	0.755
Male sex[1]	205 (56.79)	93 (43.26)	0.001	81 (48.50)	81 (48.50)	1.000
Age at randomization, years[2]	0.76 (0.22-4.1)	NA	NA	2.11 (0.34-4.77)	NA	NA
Type of congenital heart disease[1]						
Obstructive pathology	79 (21.88)	NA	NA	37 (22.15)	NA	NA
Left-right shunt	121 (33.51)	NA	NA	60 (35.92)	NA	NA
Cyanotic & not-univentricular	122 (33.79)	NA	NA	59 (35.32)	NA	NA
Cyanotic & univentricular	70 (19.39)	NA	NA	27 (16.16)	NA	NA
Other	11 (3.04)	NA	NA	5 (2.99)	NA	NA
RACHS-classification[1]						
1 or 2	168 (47.72)	NA	NA	79 (48.17)	NA	NA
3 or 4	175 (49.71)	NA	NA	84 (51.21)	NA	NA
6	9 (2.55)	NA	NA	1 (0.60)	NA	NA
Syndrome, at randomization[1]	69 (19.11)	16 (7.44)	<0.001	20 (11.98)	16 (9.58)	0.480
PELOD first 24h in ICU[2]	11 (2-12)	NA	NA	11 (2-12)	NA	NA
Socio-economic status score†	35 (24-48.5)	42.5 (29-54)	<0.001	39.5 (29-50)	39.5 (29-52.5)	0.455
At follow-up[2]						
Height, cm	107 (103-126)	122 (108-151)	<0.001	116 (104-131)	117 (107-139)	0.069
Weight, kg	18 (15-24)	22 (18-40)	<0.001	20 (16-29)	21 (18-34)	0.072
Head circumference, cm	50.7 (49.2-52.5)	52 (50.8-54)	<0.001	51 (49.5-53)	51.8 (50.5-53.4)	0.001
Age, y	4.67 (4.14-7.93)	6.75 (4.68-11.56)	<0.001	6.02 (4.21-8.78)	5.91 (4.58-9.07)	0.359

[1]numbers and percentages; [2]median (interquartile range, IQR)

Abbreviations: NA = not applicable; RACHS = Risk adjustment for congenital heart surgery[42]; PELOD = pediatric logistic organ dysfunction[43]

Table 3: Results of neurocognitive test battery in tested and propensity score matched post-heart surgery population and healthy controls in the LGC-trial

	Tested population			Propensity score matched population		
	Tested-Post-Cardiac surgery population (N=361)	Healthy controls (N=215)	p-value	Tested-Post-Cardiac surgery population (N=167)	Healthy controls (N=167)	p-value
Clinical neurological evaluation score (range, 0-8)[1]	1 (0-2)	0 (0-1)	**<0.001**	1 (0-2)	0 (0-1)	**<0.001**
Intelligence (range of possible scores, 45-155)[2]						
-Full-Scale IQ	90 (75-100)	103 (91-111)	**<0.001**	92 (78-103)	101 (90-110)	**<0.001**
-Verbal IQ	90.5 (76.2-101)	102 (92-114)	**<0.001**	94 (79.5-104)	101 (91-112)	**<0.001**
-Performance IQ	89 (77-101)	103 (92-112)	**<0.001**	91 (77-102)	101 (89-112)	**<0.001**
Visual-Motor Integration (range, 0.9-20)[2]	9 (7-10)	10 (8-12)	**<0.001**	9 (7-10)	10 (9-12)	**<0.001**
Attention, motor coordination and executive functions						
-Alertness[1]						
Reaction time dominant hand, msec	691 (447-982)	481 (320-700)	**<0.001**	566 (374-829)	544 (365-749)	0.345
Within-patient SD of repeated tests	404 (143-642)	165 (83-383)	**<0.001**	228 (115-535)	190 (98-433)	0.075
Reaction time nondominant hand, msec	697 (436-968)	499 (326-721)	**<0.001**	543 (374-791)	542 (375-744)	0.677
Within-patient SD of repeated tests	336 (158-611)	192 (87-379)	**<0.001**	221 (110-482)	216 (105-412)	0.410
-Motor coordination (No. of taps in 10s)[2]						
No. of unimanual taps dominant hand	28 (22-38)	35 (25-46)	**<0.001**	31 (22-42)	32 (24-43)	0.346
No. of unimanual taps nondominant hand	23 (17-33)	29 (21-43)	**<0.001**	25 (18-35)	27 (20-38)	0.117
No. of valid alternating taps	8 (2-18)	13 (5-30)	**<0.001**	8 (2-20)	11 (5-26)	**0.039**
No. of valid synchronous taps	16 (8-24)	21 (12-31)	**<0.001**	18 (11-27)	19 (11-27)	0.768
-Inhibition and flexibility[1]						
Δ Reaction time (inhibition), msec	313 (120-536)	200 (79-485)	**0.008**	261 (98.7-438)	258 (94-500)	0.876
Δ No. of errors (inhibition)	2 (0-3)	1 (0-2)	**0.002**	1.5 (0-3)	1 (0-2)	**0.038**
Δ Reaction time (flexibility), msec	637 (367-878)	550 (283-798)	**0.043**	603 (330-848)	623 (345-869)	0.726
Δ No. of errors (flexibility)	2 (0-4)	1 (0-3)	0.243	2 (0-4)	1 (0-3)	0.104

Table 3: Results of neurocognitive test battery in tested and propensity score matched post-heart surgery population and healthy controls in the LGC trial (continued)

	Tested population			Propensity score matched population		
	Tested-Post-Cardiac surgery population (N=361)	Healthy controls (N=215)	p-value	Tested-Post-Cardiac surgery population (N=167)	Healthy controls (N=167)	p-value
Memory[2]						
-Verbal-auditory						
Numbers (range, 1-19): Memory span (forward)	8 (5-9)	9 (7-11)	**<0.001**	8 (6-11)	9 (7-11)	0.149
Numbers (range, 1-19): Working memory (backward)	9 (6-11)	10 (8-13)	**<0.001**	9.5 (6.2-12)	10.5 (9-13)	**0.003**
Word pairs (proportion of correct responses): Learning	0.45 (0.33-0.53)	0.5 (0.38-0.66)	**0.001**	0.46 (0.36-0.57)	0.46 (0.35-0.6)	0.627
Word pairs (proportion of correct responses): Immediate memory	0.4 (0.2-0.5)	0.5 (0.35-0.64)	**<0.001**	0.4 (0.23-0.5)	0.5 (0.3-0.6)	**0.012**
Word pairs (proportion of correct responses): Delayed memory	0.3 (0.2-0.4)	0.4 (0.3-0.5)	**<0.001**	0.35 (0.21-0.42)	0.4 (0.28-0.5)	0.068
Word pairs (proportion of correct responses): Recognition	0.96 (0.9-1)	1 (0.95-1)	**0.011**	0.97 (0.9-1)	0.97 (0.93-1)	0.404
-Nonverbal, visual-spatial (proportion of correct responses)						
Pictures: Memory span	0.83 (0.71-0.89)	0.89 (0.8-0.93)	**<0.001**	0.84 (0.73-0.9)	0.86 (0.76-0.93)	0.114
Dots: learning	0.83 (0.66-0.88)	0.88 (0.83-0.94)	**<0.001**	0.87 (0.70-0.94)	0.88 (0.80-0.94)	0.061
Dots: Immediate memory	0.83 (0.62-1)	1 (0.75-1)	**<0.001**	0.87 (0.66-1)	1 (0.68-1)	0.224
Dots: delayed memory	0.83 (0.5-1)	1 (0.75-1)	**<0.001**	0.83 (0.66-1)	0.87 (0.68-1)	0.268
Learning index (range, 50-150)	93 (84-103)	101 (90-109)	**<0.001**	96 (87-103)	99 (87-109)	0.102
Behavior (by proxy), T-score[1]						
CBCL-internalizing problems (range, 29-100)	52 (45-61)	48 (41-57)	**<0.001**	51 (43-59)	49 (41-58)	**0.046**
CBCL-externalizing problems (range, 28-100)	50 (43-56)	46 (40-55)	**0.006**	49 (41-55)	47 (40-56)	0.221
CBCL-total problems (range, 24-100)	52 (45-59)	47 (40-55)	**<0.001**	51 (43-56)	48 (40-56)	**0.049**

median (interquartile range, IQR) are reported; [1]Higher scores reflect worse performance; [2]Higher scores reflect better performance

Abbreviations: SD = standard deviation; CBCL = Child Behavior Checklist

Legends of figures

Figure 1. Flow diagram of Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA)

[1]Search strategy in eMethods 1

[2]Manuscripts could be excluded for more than one reason

Figure 2. Forest plot of intelligence (author/year/journal – CHD-type)

Abbreviations: TGA = d-transposition of the great arteries; VSD = ventricular septal defect; TOF = tetralogy of Fallot; ASD = atrial septal defect; ECMO = extracorporeal membrane oxygenation; V.d.Rijken = Van der Rijken; cath = catheterization; surg = surgery; CA = circulatory arrest; CPB = cardiopulmonary bypass; CI = confidence interval

Figure 3. Forest plot of attention (Att) (alertness) non-reaction time (author/year/journal – CHD-type / attention measure)

Legend: Att1 = Mean accuracy (%) Attention Network Test; Att2 = Visual Attention (NEuroPSYchological Assessment, NEPSY); Att3 = Auditory Attention (NEPSY); Att4 = Processing Speed (Wechsler Intelligence Scale for Children, WISC-IV); Att5 = Delay Task, Vigilance hits; Att6 = speed of information (BAS, British Ability Scales)

Figure 4. Forest plot of verbal memory (Mem) (author/year/journal – CHD-type / Memory measure)

Abbreviations: Mem1 = Sum of Immediate and Delayed Verbal Memory (Children's Memory Scale, Word Pairs, proportion correct responses); Mem2 = Narrative Memory (NEPSY); Mem3 = Verbal Memory (Wide Range Assessment of Memory and Learning, WRAML)

Figure 5. Forest plot of executive function (EF) (inhibition) reaction time (RT) (author/year/journal – CHD-type / EF measure)

Abbreviations: EF-Test1 = Incongruent Stroop; EF-test2 = Stroop; EF-test3 = Conflict (Attention Network Test); EF-test4 = Inhibition (Amsterdam Neuropsychological Tasks)